DOCUMENT RESUME

ED 371 812 JC 940 421

AUTHOR Leitzel, Thomas C.; Vogler, Daniel E.
TITLE Curriculum Alignment: Theory to Practice.
PUB DATE May 94
NOTE 51p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Community Colleges; *Course Evaluation; *Curriculum Development; *Evaluation Methods; Inservice Teacher Education; *Instructional Development; *Student Evaluation; *Teaching Methods; Two Year Colleges
IDENTIFIERS *Curriculum Alignment

ABSTRACT

Curriculum alignment is the conscious congruence of three educational elements: curriculum, instruction, and assessment. Alignment is rooted in the belief that instructional plans are established through outcomes-based content goals and the goal of assuring that delivery and assessment are congruent. Platform unity, based on the Principles of Performance Instruction, is a way to uphold curriculum alignment. The integration of planning and evaluating is often neglected in traditional approaches to instruction. Performance Instruction holds that course content should be planned, delivered, and evaluated consistently to assure unity. Test creation, for example, needs to be related to content planning decisions. The domain-level at which content is planned becomes the basis for creating test items, with content planning and testing at the same domain level to assure unity. In the theoretical literature, considerable attention is given to faculty's inability to plan and test content consistently. These are arguably important and necessary faculty tasks, and should be the focus of staff development. Unfortunately, even the literature on planning and testing treats them as separate, independent activities. Many faculty have had no formal coursework or in-service training in assessment. As a result, classroom tests are usually short, objective, and of poor technical quality and usually call for the memorization of facts. There is little research in the assessment field regarding practical tools to help faculty evaluate criterion-referenced tests. By looking at the theoretical literature on planning and evaluation, one sees the need to move to practical implications for the benefit of aligning the planned and tested curricula, thus achieving efficiency, effectiveness, and overall unity of instruction. (Contains 104 references.) (KP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 371 812

CURRICULUM ALIGNMENT: THEORY TO PRACTICE

by

Thomas C. Leitzel
Dean of Business, Health and Technologies
Central Piedmont Community College
Charlotte, North Carolina

and

Daniel E. Vogler
Associate Professor
Community College Program Area
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T. Leitzel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

May, 1994

BEST COPY AVAILABLE

JC 940 421

Table of Contents

Overview 1

Part I: Curriculum Alignment Theory:

The Basis for Alignment 4

Alignment Terminology 4

Curriculum Alignment Implications 6

Platform Unity 7

Misalignment 7

Planning 8

Taxonomy of Educational Objectives 8

Criticisms of the Classification Scheme 10

The Cognitive Domain 11

Goals. 12

Objectives 13

Stating Instructional Objectives 15

Simplifying Faculty Course Planning 16

Testing 17

Test Categories 18

Teacher-Constructed Tests 20

Classroom Testing 23

Purposes of Classroom Tests 24

Planning the Test 26

Integrating Planning and Testing 27

Selecting Appropriate Test Types 28

Part II: Curriculum Alignment Practice:

Principles of Performance Instruction 30

Content Goals 32

Building an Exam 34

Table of Test Type Specifications 35

Summary 37

References 40

CURRICULUM ALIGNMENT: THEORY TO PRACTICE

Overview

The most fundamental responsibility of an instructor is to determine what should be learned, and therefore, taught in a course. This responsibility is moot, however, if the instructor fails to deliver or evaluate what was planned. Planning, delivery, and evaluation are three elements of instruction. The situation that exists when planned content is delivered and then evaluated is alignment. Curriculum alignment, according to Savard and Cotton (1982), is a term used to denote the conscious congruence of three educational elements: curriculum, instruction, and assessment. This suggests that incongruence occurs when faculty fail to deliver or evaluate planned content. In a recent study in community colleges, institutions heralded for their attention to teaching, Leitzel and Vogler (1993) analyzed alignment and found evidence to suggest that faculty are uninformed about this fundamental principle.

Curriculum alignment is not a well recognized term in the literature. An ERIC search revealed only two citations when "curriculum-alignment" was cross referenced with "research" from a 1966 - 1981 ERIC database and 11 citations using a 1982 - December, 1993 ERIC database. Despite the slight increase of citations, Crowell and Tissott (1986) rightfully observed a general lack of attention to alignment. They concluded, very little practical research exists to guide the efforts of schools and advocated the implementation of practical procedures to aid in understanding and determining alignment.

Most references to alignment in the literature focused on alignment in secondary schools. The most frequent references to alignment suggest it is used as a measure to compare curriculum decisions created at a policy-making level with the actual content delivered in the schools or with determining how well textbooks align with planned content. By way of contrast, an ERIC search using the 1982 - December, 1993 database on the word "pedagogy" yielded 1,492 hits; "curriculum" 50,477 hits; "instruction and planning" 5,780 hits; "instruction and delivery" 1,471 hits; and, "instruction and evaluation" 15,953 hits.

Alignment can be analyzed in three ways: planning to delivery (PD), delivery to evaluation (DE), and planning to evaluation (PE). This concept is illustrated in Figure 1 through a triad arrangement of the three elements of instruction.

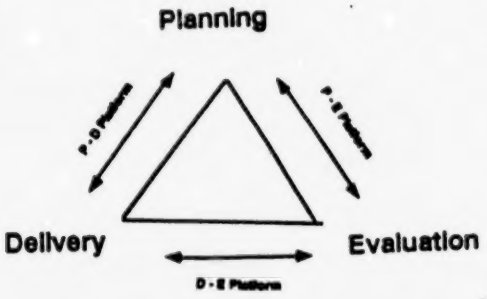

Figure 1

The Platforms of Alignment

This paper will track the progression of the literature on alignment from theory to practice. The paper will accent the relevant information needed to assess the planning/evaluation (PE) platform for cognitive content. Application of the concept to

include delivery as associated with planning and evaluation in any domain is not the purview of this paper. This delimitation is prompted by logic considerations that suggest the cognitive domain and the PE platform are the prudent parameters for analysis of alignment.

The topic of analyzing alignment via planning and evaluation represents a challenge, albeit one that is possible. It is a reasonable place to begin a study on alignment. On the other hand, it is more difficult to analyze alignment using the delivery construct. While planning and evaluation tend to have documented information, delivery is instructor controlled. As it is often said, once the classroom door closes, only the instructor controls what happens. It is not impossible to measure alignment using the delivery construct, it is just more difficult. There is definitely a study in the waiting if one were to use the delivery construct in an alignment analysis. For now, we believe we have only begun to influence discussion on the importance of alignment through a numerical interpretation. As it is presented here, alignment is manifested through Performance Instruction, a model conforming to the congruence of planning, delivery and evaluation.

Part I: Curriculum Alignment Theory

This section concentrates on the theory of alignment by providing a point of reference structured around terminology. Several definitions, such as alignment, platform unity, and misalignment, are offered to establish a basis which is fundamental to the overall concept of curriculum alignment. Instructional elements including planning and testing are detailed.

The Basis for Alignment

Alignment Terminology

Various terms are used in the literature to describe the condition whereby three components of the curriculum, the written, the delivered, and the evaluated, have explicit relationships to one another. Among these terms are *congruence* (Brickell, 1976), *overlap* (Leinhardt & Seewald, 1981), *curriculum test overlap* (Hartzell, 1984), *match, linking* (Airasian & Madaus, 1983), and *alignment* (English, 1986-1987). Considering these terms, curriculum alignment, or just simply alignment, is most closely related to the purpose of this paper.

A notable advocate of curriculum alignment has been Fenwick W. English. For English, alignment consists of the written, the taught, and the tested curriculum. In his National Academy for School Executives, English (1992a) developed a thorough inquiry into alignment elements. Among terms he used to describe the written element are *curriculum* and *the work plan*. Terms used to describe the taught element include *delivered* and *the work*. The

tested element included such terms as *the measured, evaluated,* and *the work measurement.*

In developing an all-encompassing definition Crowell and Tissot (1986) refer to alignment as the congruence of all elements of a school curriculum, including the curriculum goals, the instructional program, and the tests used to judge outcomes. Savard and Cotton (1982) view alignment as the congruence between state curriculum guidelines and locally-developed instructional practices.

Similar to Figure 1, English (1979, 1986-1987, 1988, 1992a, 1992b) used a triangle to depict this concept and it became the universal symbol to represent alignment in the literature. Savard and Cotton (1982) referred to it as the *alignment triangle*, with objectives, instruction, and testing anchoring the corners. Relating back to the descriptors used in Figure 1, objectives refer to the written or planned curriculum, instruction the delivered curriculum, and testing refers to the evaluated curriculum.

Although English's concept of alignment was developed primarily for use in secondary school management through a process he termed an audit (English, 1988), it is generic to instruction at all levels. For instance, Pautler (1989, 1990, 1992) advocated using alignment in community colleges. He believed that all institutions, regardless of level, are concerned with the improvement of instruction. According to Pautler:

> "curriculum alignment should be in the vocabulary of community college faculty and administrators. In this day and age of so many part-time instructors being

> employed by community colleges, the need for alignment is even more critical. The process of alignment is directly related to the development and use of a functional course of study" (1989, p. 178).

English (1992a) distinguished two types of curriculum alignment, design alignment and delivery alignment. Design alignment is the relationship between the curriculum and the test. Delivery alignment is the relationship of what is taught to the test and to the curriculum.

Curriculum Alignment Implications

English (1987) believed curriculum alignment begins with the written curriculum. English and Steffy (1983) called the written curriculum a *prescription* or a *set of specifications*. Ideally, planning comes first in a sequence of events that also includes delivery and evaluation.

English and Steffy (1983) observed that written curriculum problems are usually design in nature that stem from errors made in planning. For example, a design problem occurs when what is planned is not linked to test construction. Consequently, students may be tested on skills or knowledge for which they have not been taught. A faulty design, no matter how capably or uniformly implemented, prevents optimization until it is corrected so that plans and tests are congruent (English & Steffy, 1983).

Pautler (1990) believed that developing written curriculum documentation of courses of study may be one of the least understood activities within community colleges. According to Pautler (1990), most college-level courses lack a consistent

format. The design of functional courses of study parallels curriculum alignment as a worthwhile effort to improve instructional effectiveness of the college (Pautler, 1990). Pautler (1990) advocated instructional design principles to develop courses of study. He believed if community college staff were given more preparation in instructional design principles, then improved institutional effectiveness would result. Accordingly, alignment should be an explicit part of the faculty development process and be implemented in the instructional design process (Pautler, 1990).

Platform Unity

Figure 1 showed the three elements of curriculum at three points in a triangle. If the three elements of the curriculum were placed at the three points on a triangle, along one plane, or platform of the triangle, is the relationship between the planned and tested curriculum; hence the origination of the term, "platform unity." Figure 1 depicts this relationship along the P-E Platform. This term was selected to simplify the analysis of these two constructs. By definition, platform unity is a curriculum-based estimate of congruence between planning and testing.

Misalignment

Misalignment is a condition that occurs when content plans and testing decisions are not congruent. Another term for this is *disunity*. The importance of curriculum alignment is accented when misalignment occurs. Most students can tell stories about instructors whose lectures did not follow the course syllabus and

whose final exam had nothing to do with either (Pautler, 1989). Nitko (1989) stated that tests created by classroom teachers often inadequately measure the teachers' espoused instructional goals. When misalignment occurs, the written, taught, and tested curriculum function independently. The result is that what faculty plan exists apart from what students encounter on tests (Savard and Cotton, 1982). According to Nitko (1989), there are several undesirable consequences of using tests that are inadequately linked to instructional plans. These include the following four:

> "(a) teachers and students might be inappropriately informed about students' learning progress and learning difficulties;
>
> (b) students' motivation for learning could be reduced;
>
> (c) critical decisions about students (e.g., whether to award a student a degree) might be made unfairly; and
>
> (d) the effectiveness of instruction may be evaluated incorrectly" (p. 447).

Systematic planning, delivery and evaluation of course content is a way to support the principles of curriculum alignment.

Planning

Planning instruction is a necessary faculty activity, though it is rarely given proper attention. According to Cross (1987), the processes involved in classroom teaching in higher education should be taken much more seriously than they have been in the past. Recent advances in cognitive psychology indicate that students learn better if teachers clearly communicate course goals,

objectives, and discipline structure (Stark, Lowther, Sossen & Shaw, 1991).

Taxonomy of Educational Objectives

A step toward recognizing the complexity of educational objectives and the difficulties involved in measuring student achievement was taken when Bloom, Englehart, Furst, Hill, and Krathwohl (1956) worked to develop taxonomies of educational objectives. The main reason for developing the taxonomies was to facilitate communication among educational researcher, curriculum developer, and evaluator (Payne, 1992). The taxonomy matched educational goals and test items in the cognitive domain.

According to Bloom, Englehart, Furst, Hill, and Krathwohl (1956), learning occurs in three different domains: cognitive, psychomotor, and affective. There are six major classifications of this taxonomy, arranged in hierarchical order from simple to complex. The classifications include: (1) knowledge of factual information; (2) comprehension of knowledge; (3) application of knowledge and understanding to specific situations; (4) analysis of complex ideas into their component parts; (5) synthesis of disparate items into a whole; and (6) evaluation of the consequences of actions. Payne (1992) promoted the taxonomy's usefulness for analyzing instruction, as in comparing the emphases in course planning with those in test questions.

Criticisms of the Classification Scheme

Though Bloom's Taxonomy is certainly a noteworthy contribution to educational classification, it is not without its problems. Researchers generally support the cognitive domain taxonomy. Support for the order of the more complex categories has largely failed to develop. According to Stahl and Murphy (1981), the taxonomy has been and is being used even though it is not consistent with any presently accepted theory, model, or approach to human memory, thinking, or learning. Several researchers (Kropp, Stoker & Bashaw, 1966; Metfessel, Michael, & Kirsner, 1969; Stanley & Bolton, 1957) have reported that persons frequently disagree on the taxonomy level represented by many items, except those at the knowledge level.

In the taxonomy, the same term may be used at several levels. The verb "identify", for example, appears at the knowledge, understanding, application, and analysis levels. Although this may be appropriate, the overlap of terms causes confusion. Diamond (1989) believed faculty were turned off because many of the early advocates of stating goals in performance terms focused on minutiae and on complex classification systems that faculty did not understand. According to Diamond (1989), faculty spend a great deal of time analyzing the type or level of their objectives. He believed it is far more essential that efforts are made to ensure that useful statements be written, that they include all of the elements that should be addressed, and that they be measurable within the context of the course. No matter how the objectives are

conceived or categorized, if they are clearly stated, then it is easier for the instructor to plan content, deliver content, and to prepare tests to assess the extent to which objectives are achieved (Diamond, 1989).

The Cognitive Domain

The cognitive domain, which includes instructional objectives that deal with recall or recognition of learned material and the development of intellectual abilities, is at the core of curriculum test development (Stanley and Hopkins, 1972). The largest portion of educational objectives fall into this domain (Gaff, 1975; Krathwohl, et al., 1964; Payne, 1992).

Moving up the hierarchy, each category is assumed to include behaviors at the lower level; higher-level skills utilize lower-level ones and demand greater intellect from the learner (Jacobs & Chase, 1992). For example, comprehension includes knowledge because students must have to comprehend material before they can apply or analyze it. The application level includes knowledge and comprehension, and so on. Evaluation is the most demanding of the cognitive skills because it requires judgments from all previous levels.

Two other domains in the taxonomy include the affective and psychomotor. The affective domain includes those objectives that emphasize feelings and emotions, such as interests, attitudes, appreciation, and methods of adjustment. The psychomotor domain includes those objectives that emphasize motor skill, such as

handwriting, typing, swimming, and operating machinery (Gronlund, 1985).

Goals

Goals are the philosophic intentions of education. Establishing goals for instruction is a common method of planning. Goal-referenced instructional models focus on desired, observable learner behavior produced as the result of instruction (Baker & Popham, 1970). Baker and Popham frame the basic planning decision. They suggest, instead of asking what shall I do, the instructor needs to ask, what do I want my learners to become. By introducing this question, the emphasis shifts to a student-oriented planning model. There is a shift occurring in education away from course-centered and content-centered approaches of course development to more learner-centered approaches (Menges & Mathis, 1988).

Diamond (1989) believed the process of determining whether or not academic programs are successful begins by stating what is expected of students. This requires describing goals and objectives in performance terms and creating evaluation instruments that assess the abilities of students to meet specific criteria. Developing a course plan that optimizes each aspect of instruction is a challenging faculty activity requiring expertise and informed decision-making.

Educational goals must be stated in more precise and observable form to give direction to the important tasks of curriculum development and evaluation (Stanley & Hopkins, 1972). According to Cook (1978), a clear, useful, verifiable statement of

student learning outcomes should contain at least three ingredients. First, it should specify the learner; second, it should describe an unambiguous and observable action; third, it should specify the conditions under which these outcomes will be assessed. Others believe it should contain a description of the minimal level of acceptable response (Gagne, 1975; Gerhard, 1971; Mager, 1973; Popham, 1973).

Objectives

While goals communicate curricular intent, objectives are communication devices for students (Vogler, 1991). Objectives provide instructors with a method of specifying instructional goals in a curriculum. They are, by definition, intended outcomes of instruction. At the course level, objectives provide a framework and a guide for the specific instructional decisions that follow (Gunter, Estes & Schwab, 1990). Vogler (1991) stated that "objectives describe student outcomes, the standards used to measure outcomes, and the conditions under which the outcomes will be produced" (p. 29). The terms "behavioral", "performance", and "measurable" objectives have been used interchangeably to denote the end behaviors desired as a result of instruction (Gaff, 1975). Course objectives written in behavioral, observable and measurable terms state precisely what students will be able to do as they interact with specific course content; they specify what students will do or produce when the instructional goal has been met or after completing the instructional unit (Bloom, Madaus & Hastings, 1981; Copperud, 1979; Gerhard, 1971). These statements also

describe how behavior will be measured (Gerhard, 1971). In the planning process, faculty take their plans and make a communication statement to students, usually in the form of a course syllabus, which contains the course objectives. The syllabus paves the way to student acceptance and understanding of the instructional system (Kibler, Cegala, Watson, Barker, & Miles, 1981).

There has been considerable professional attention directed toward the nature of objectives, the way they should be formulated, and the functions they serve in the design of instructional sequences (Baker & Popham, 1970; Diamond, 1989; Gunter, Estes & Schwab, 1990; Kibler, Cegala, Watson, Barker, & Miles, 1981; Mager, 1962, 1973; Popham, 1973, 1978; Salvia & Hughes, 1990; Vogler 1991). Although the professional literature recommended that objectives should describe precisely what is expected of students, Diamond (1989) believed most faculty conceptualize their teaching in terms of their content area rather than with reference to student outcomes. Although they strive to plan and teach courses in ways that help students learn effectively, faculty members seldom have received specific training for these tasks (Stark, Lowther, Sossen, & Shaw, 1991).

When planning a course, it is necessary to plan instructional outcomes, not instructional activities. Consequently, it is important to focus planning statements on student performance from a student's point of view rather than on teacher performance (Feldhausen, 1980; Salvia & Hughes, 1990). For example, reading Plato's Republic and discussing portions of it in class describes

a planned activity. It does not indicate what students are supposed to gain from the activity. It is more appropriate to state the learning task in behavioral terms, such as the identification of major features of Plato's theory.

Mager (1975) and Diamond (1989) posited that a complete objective has three elements: (a) an action or behavior that the student is to perform; (b) the conditions under which these actions are to be performed; and (c) the criteria or standards for saying whether these actions have been performed satisfactorily. The most important element is the behavior (Mager, 1975).

Stating Instructional Objectives

In stating objectives, teachers "seek to clarify in their own mind and to communicate to others the sought-for change in thoughts, actions, or feelings that a particular unit or educational program should help a student realize" (Bloom, Madaus & Hastings, 1981, p. 17). In preparing a list of instructional objectives for a course of study, there are two immediate goals. One is to obtain as complete a list of objectives as possible. The other goal is to state objectives so that "they clearly indicate the learning outcomes that one expects from instruction" (Gronlund, 1981, p. 45). Statements of specific learning clearly convey instructional intent if each statement begins with an action verb that indicates definite observable responses (Copperud, 1979; Gronlund, 1981, 1985; Tuckman, 1975; Vogler, 1991). Verbs that have discrete, definite meanings should be used rather than those open to many interpretations (Stanley & Hopkins, 1972). For

example, the words "to identify", "to solve", or "to construct" are preferable to "to understand" and "to appreciate" (Mager, 1962).

The first step in a systematic approach to instruction is specifying objectives in operational terms (Baker & Popham, 1970). To formulate useful statements, educators must choose words that are interpreted in the same sense by readers (Bloom, Madaus & Hastings, 1981). Objectives that are tied to performance levels provide a basis for evaluating a range of performance (Diamond, 1989). English and Steffy (1983) stated that objectives used in planning instruction should focus on the knowledge, skills and attitudes a student should master, lend themselves to measurement, and reflect research knowledge about the subject.

Cognitive objectives specify that learners will do something with knowledge. Gunter, Estes, and Schwab (1990) specified two different kinds of cognitive objectives, declarative and procedural. Declarative objectives represent knowledge that can be expressed in true or false statements. Procedural knowledge results from the reasoning process. The most common instructional objectives in postsecondary education are declarative, that is, they focus on information recall (Babbin, 1987).

Simplifying Faculty Course Planning

After studying over eighty instructional projects, Bergquist and Armstrong (1986) reported that the inability to state course goals in performance terms was a major problem. According to Diamond (1989), faculty resent stating instruction in behavioral terms and deem the task a waste of time. Consequently, they write

objectives at a trivial level, which are probably the easiest to write.

Gaff (1975) believed that by stating objectives with a high degree of specificity, it is possible for everyone involved to recognize the behavior that meets or fails to meet them when evaluating. With this structure, the specific objectives inform instructors what their tests should be like. They tell what will be covered and with what emphasis (Jacobs & Chase, 1992). Considering the importance of planning and evaluation, Greeno (1976) believed the development of instructional objectives begins with considering the kinds of tests used to assess whether students have acquired the knowledge intended as the outcome of learning.

Testing

Tests are an integral part of college courses, typically constructed to measure whether objectives have been met (Tuckman, 1975). Another purpose is to obtain valid, reliable, and useful information concerning student achievement so that learning can be improved (Gronlund, 1982). The instructor attempts to determine if students can behave as intended when the instruction was planned.

Lozak (1987) estimated that between one fourth to one third of an instructor's time is spent on measurement. According to Stanley and Hopkins (1972), constructing a test is one of the hardest jobs a teacher has to perform. It demands an understanding of the objectives being assessed and of the examinees and their test-taking behavior. In constructing classroom tests, a major concern

is that test items call for particular types of performance indicated in the specific learning outcomes pertinent to each instructional objective or content goal (Gentry, 1989; Gronlund, 1988; Heywood, 1989; Popham, 1978, Vogler, 1991). Baker and Popham (1970) stated that objectives and evaluation should be identical; test items should be drawn from the class of behavior specified in the objectives.

Test Categories

Tests are either criterion-referenced or norm-referenced. Criterion-referenced exams compare an individual's score to some specified level of performance or criteria; norm-referenced exams compare the individual's score to those of individuals in a group (Jacobs & Chase, 1992). In criterion-referenced testing situations, the meaningfulness of an individual score does not depend on a comparison with other testees. One seeks to know what the individual can do, not how he or she stands in comparison to others. When testing to a criterion, the score indicates within error limits exactly what an individual can and cannot do (Popham, 1973).

With criterion-referenced tests, scores are compared with specific standards. In criterion-referenced testing situations, instructors are not concerned with how difficult the items are or how well the items discriminate between students. The key element is how well the items reflect the specific learning tasks (Jacobs & Chase, 1992). While Tuckman (1985) indicated that criterion-referenced tests in education are used to monitor student progress

and evaluate instructional program effectiveness, Popham (1973) argued that few analyses have ever been made of the practical implications of these tests.

Typical norm-referenced measures, such as item-analysis procedures and internal-consistency estimates, as well as the standard notions of validity and reliability are designed to facilitate comparisons among individuals. It is not possible to tell a norm-referenced test from a criterion-referenced test by looking at it. Although a criterion-referenced test could also be used as a norm-referenced test, the reverse is not easy to accomplish (Popham, 1973).

The contrast between norm-referenced and criterion-referenced measurement demonstrates how current test construction and test improvement has purposes other than measuring the quality of instructional programs. "Mechanisms such as item-analysis procedures and internal-consistency estimates, as well as the standard notions of validity and reliability are not pertinent to the kinds of measurement procedures that must be used to assess the attainment of measurable instructional objectives" (Popham, 1973, p. 7-8).

Popham (1973) believed a great deal of sophisticated work needs to be done on how to devise measures that satisfactorily serve to sample the behavior and content domains delimited by a well-stated objective. To go along with this, Popham suggested using efficient ways to produce satisfactory measures of criterion-referenced objectives so that they can be economically assessed.

Teacher-Constructed Tests

A measure more closely aligned with taught content is the teacher-developed test. Ebel (1965) believed classroom tests prepared by the faculty member are likely to fit content and objectives of a particular course better than a test prepared by anyone else. MacCuish (1986) reported that 91 percent of college faculty create their own tests. However, the process of constructing a good test item is time consuming. It demands an understanding of the objectives being assessed and of the examinees and their test-taking behavior (Stanley & Hopkins, 1972).

Tests differ in what type of interpretations are made from the scores, how the content of the test is developed, and the appropriateness of the tests for making various educational decisions (Salvia & Hughes, 1990). Publisher-prepared tests include items that have not been carefully prepared or critically reviewed by other experts in educational measurement or in the subject field (Ebel, 1965). Further, the content of published tests often does not match the content that is taught. While these tests may reflect students' intellectual abilities, they are useless for evaluating what a student has learned from school instruction. Unfortunately, some instructors, outstanding in their scholarship and teaching ability, possess naive notions about the requirements for effective measurement of educational achievement (Ebel, 1965). Findings suggest teaching and testing often occur to a considerable extent at the lowest level (Stanley & Hopkins, 1972; Ball, Doss & Dewalt, 1986). Students are not

encouraged to use their minds at higher levels. One of the most common defects of teacher-constructed tests is the lack of relevant, difficult items. Errors that occur in testing include writing too many questions for a certain topic, including only a few questions from a certain part of the course, and completely ignoring another part of the course (Copperud, 1979).

Some faculty are more inclined to ask about the specific, incidental details than about the important general principles (Ebel, 1965). Jacobs and Chase (1992) reported that tests:

> "include too many questions requiring only knowledge of facts and trivia and provide no intellectual challenge to the students. Students often complain that test content does not reflect the material discussed in class or what the professor seemed to indicate was most important. They often feel somewhat cheated when they have put forth a great deal of effort and then take a test that does not permit them to show what they have learned" (p. 13).

Teachers display weaknesses in test construction. "Content-oriented teachers can become so preoccupied with the importance of students' learning specific facts and terms that their students fail to gain a sense of direction and a level of understanding, which comes from seeing the interrelatedness of concepts and principles and applying them to unfamiliar problems" (Adams, 1966, p. 329).

Fleming and Chambers (1983) analyzed over 400 teacher-developed tests and thousands of test items. They found that (a) teachers use short answer questions most often; (b) teachers avoid essay questions; and (c) teachers used more matching items than multiple-choice items. They also found that teachers devise more

test questions to sample knowledge of facts than any of the other behavioral categories; teachers develop few questions to test application behaviors.

Proper balance improves test reliability and validity (Gentry, 1989). It also lessens the chance that the test will be filled with items dealing with the instructor's "pet topics" or items easy to construct. It also makes it less probable that topics or content will be ignored. Stanley and Hopkins (1972) recommended using a table of specifications to guide the test maker so that the test essentially mirrors the instructional objectives.

Most college instructors feel poorly prepared to construct tests in their classes because they have never received any kind of formal training in this area (Jacobs & Chase, 1992). Lozak (1987) blamed the higher education system for not preparing its graduates for the assessor role. Goslin (1967) found only 40 percent of elementary and secondary teachers had as little as one class in testing and measurement. Yeh (1978), in a study of teacher use of test results, reported that only 50 percent of the teachers sampled were able to correctly interpret percentile ranks and grade equivalents. Consequently faculty are uncomfortable with test construction and view it as a difficult chore. Milton and Associates said, "testing is perhaps the most neglected feature of good instruction" (1978, p. 101). Stanley and Hopkins (1972) concluded that if evaluation procedures are poor, then the quality of the information on which judgements are based cannot be adequate.

Students study what they think they are going to be asked in the instructor's tests (Anderson, 1987). If students believe instructors will emphasize factual recall, students will learn facts. "If instructors are going to ask them to apply material to new situations, students will study and learn how to apply principles and theories. If students believe they will be required to analyze, synthesize, and make judgments, then they will learn to use higher-level thinking skills" (Jacobs & Chase, 1992, p. 5-6). Instructors need to gain a perspective on their purposes of testing since tests and examinations profoundly influence what students study (Anderson, 1987).

Classroom Testing

Testing is an integral part of instruction. Unlike the demand for expanded use of test scores for accountability, largely the creation of legislators, administrators, and policy boards, Linn (1989) believed the renewed interest in testing was to make better day-to-day instructional decisions. The movement to make better instructional decisions is spurred by educational measurement specialists and cognitive psychologists. The instructional use of test results is not to predict who will succeed or to report results to satisfy accountability demands. Rather, it is to help individual students gain the most from instruction (Linn, 1989).

According to Isaac and Michael (1990), the purpose of testing is not on theory building but on product delivery and mission accomplishment. Dyer (1970) stated that the measurement process is usually thought of only in terms of numerical test scores or other

quantitative descriptions of the phenomena being measured. Nitko (1989) observed that traditional inquiry into test design has focused on optimizing measurement efficiency rather than on optimizing instructional efficacy.

Tests given during the course assess the extent to which students are achieving the stated objectives. Jacobs and Chase (1992) believed:

> "classroom tests provide feedback on what students are learning in the class and thus help instructors to determine the effectiveness of their teaching of a particular segment of the course. By analyzing the test results, faculty can identify problem areas that perhaps should be reviewed or retaught before students move on to new material" (p. 3).

The idea that testing should help students learn better is not really new. Cook (1951) concluded that all the functions of educational measurement are concerned either directly or indirectly with the facilitation of learning. Tyler (1951) argued that this facilitation requires linking testing and instruction. He viewed educational measurement not as a separate process, but an integral part of instruction.

Purposes of Classroom Tests

Instructors commit a common error in testing when they fail to communicate to students the purpose of their classroom tests and how the results will aid learning (Jacobs & Chase, 1992). According to Linn (1983), there are three purposes of classroom tests: (a) to provide feedback to students and to the teacher; (b) to flag facts or concepts that are considered important; and, (c) to determine grades.

According to Adams (1966), tests provide students with tangible indicators of the outcomes expected from a course, even to a greater degree than do the textbook or syllabus. It is largely the in-class, teacher-made test that provides students with confirmation or feedback concerning the effectiveness of their efforts to learn (Adams, 1966). Feedback from tests also helps teachers provide more appropriate instructional guidance for individual students and the class as a whole. Beyond providing feedback, tests are used as a means of demonstrating mastery or certification (Fortune, 1985; Stanley & Hopkins, 1972).

A test is often given to students before admission to a program of studies to increase the quality of selection and classification (Stanley & Hopkins, 1972). If the admission policy of the institution changes, the purpose of the test can change. For example, if an admission policy were liberal and encouraged admission to a certain program, placement testing requirements might be less stringent. For such purposes, a condition of undertesting content may be legitimate.

Pretests also provide a baseline to evaluate the changes over a semester's time and hence the effectiveness of the instruction that has taken place (Jacobs & Chase, 1992). Tests at the beginning of the course serve as a pretest of content knowledge so faculty can measure students' knowledge of prerequisite material or concepts to be covered in the course and can plan their presentations accordingly (Jacobs & Chase, 1992).

When faculty review, interact with, or practice skills and

concepts after they have been mastered, they engage in what psychologists call overlearning. In certain situations, faculty may purposely want students to overlearn content. Examinations are a useful means of overlearning (Stanley & Hopkins, 1972). Overlearning contributes to long-term retention. When tests are designed for such purposes, a condition of overtesting may be legitimate.

Planning the Test

Before any test items are written, Jacobs and Chase (1992) urged instructors to develop a plan that specifies clearly what they are going to measure. A test plan enables the instructor to create an instrument that tests what has been planned and taught. Typically, a test plan has two components: the content to be covered and the cognitive skills to be measured (Jacobs & Chase, 1992). Test item content thus relates back to some plan. In the second component, instructors are concerned with measuring cognitive skills or the intellectual processes that students use to construct answers to questions.

Boersma (1967) found that teachers systematically involved in applying evaluative criteria had a clearer perception of the curriculum. Bloom (1961) noted that when teachers participated in the construction of tests in a systematic way, not only were objectives clarified, but more relevant instruction also occurred. Ebel (1965) insisted that classroom tests be prepared by the faculty member. He said, "if testing and teaching are in the hands of the same person, they are likely to be more effectively

integrated in the total educational process than if the testing were separated from the teaching" (Ebel, 1965, p. 9). To insure that the curriculum drives the tests, test construction should follow and be dependent upon curriculum development (English & Steffy, 1983). Professional development for faculty should focus on test construction and measurement techniques.

Integrating Planning and Testing

Gronlund (1985) believed systematic procedures to obtain a representative sample of student performance is the only assurance that a classroom test validly measures instructional objectives and course content. A device widely used for this purpose is the two-way, table of specifications chart. Adams (1966) believed there was sound rationale for building a table of specifications.

> "If a teacher accumulates test items without a plan, they will unduly represent informational learning, especially knowledge of specific facts. Moreover, teachers are likely to overemphasize certain areas of content in which items are easily constructed. As a way to improve the test's representativeness, or its content validity, one should first develop a blueprint for the test" (p. 327).

The chart relates instructional objectives to course content and specifies the relative emphasis to be given to each type of learning outcome. On the chart, content topics are listed on one dimension and cognitive skills on the other. According to Adams (1966), a two-way table of specifications gives adequate consideration to both the content and cognitive abilities of the course. He believed it was important that test coverage be adequate from both points of view.

To build a table of specifications one must (a) obtain a list of instructional objectives, (b) outline the course content, and (c) prepare the two-way chart. Gronlund (1981) argued that effective evaluation of student learning relates evaluation procedures as directly as possible to the intended learning outcomes. This is more easily accomplished if the instructional objectives and the specific learning outcomes have been clearly stated in performance terms. It is then a matter of constructing or selecting evaluation instruments that provide the most direct evidence concerning the attainment of the stated outcomes (Gronlund, 1981).

Selecting Appropriate Test Types

Items used in classroom tests are generally divided into two general categories: (a) structured, objective items that require students to select the correct answer from among a limited number of alternatives, and (b) subjective items that require students to provide responses (Gronlund, 1981). Powell and Gillespie (1990) used the term selected-response to refer to what Gronlund called objective test items and constructed-response for subjective test items.

Selected-response tests require much more time to create, but scoring them is relatively quick. While questions for a constructed response test are relatively easy to prepare, they are much more difficult to grade. Despite scoring complexities, the

use of the constructed-response test is rising (Powell & Gillespie, 1990).

The specific type of item selected depends on the nature of the objective to be measured (Gronlund, 1981; Stanley & Hopkins, 1972; Vogler, 1991). If faculty want students to learn facts, they test them on their ability to recall information; if they want to promote critical thinking, they write test items that emphasize understanding, application, and other higher-level skills (Jacobs & Chase, 1992). The process of relating test items as directly as possible to specific outcomes provides a greater assurance that the test is a valid measure of the instructional objectives (Gronlund, 1981).

Advantages and disadvantages of subjective and objective test items have been cited by authorities in the field of teaching and measurement (Cunningham, 1986; Ebel & Frisbie, 1986; Green, 1975; Jacobs & Chase, 1992; Mehrens & Lehmann, 1984; Payne, 1974; Roid & Haladyna, 1982; Swezey, 1981; Thorndike & Hagen, 1969; Vogler, 1991; Wesman, 1971). Each type of test item is efficient for measuring some learning outcomes and inefficient for measuring others. Popham (1978) stated that for measuring knowledge of factual information, the selected-response test is more efficient. If the aim is to measure originality, the ability to synthesize ideas, write effectively, or to solve problems, constructed-response tests are better (Powell & Gillespie, 1990).

Part II: Curriculum Alignment Practice:

A step beyond theory is practice. This section concentrates on the implementation of planning and testing theory through the use of an instructional model that conforms to the principles stated previously. The details of theory are preserved and accessed by users through expert system technology. The practical suggestions in this section reinforce the theory and serve as a mechanism for highlighting the importance to which planned and evaluated instruction have in curriculum alignment.

Principles of Performance Instruction

Performance Instruction is the planning, delivery, and evaluation of learning and teaching. It is based upon the Vogler Curriculum-Pedagogy-Assessment (C-P-A) model (Vogler, 1991). The model is the foundation of the expert system for three independent or integrated software packages. The model works without the software; however, the software guides, monitors and aggregates user decisions. The information presented below briefly explains some of the features of the expert system in the curriculum model.

The expert system in the software tutors t' : user in correct course design, lesson planning, and the creation of test items and criterion referenced examinations. It provides a standard that promotes consistency within the curriculum.

The system is dynamic by providing the infrastructure to develop functional outputs--a syllabus, a lesson plan or individualized module, or an exam, monitoring decisions made by the

user. The expert system is passive by providing a prescriptive manuscript that can be accessed by the user in hard copy or in electronic form on the screen. The software flexibility is enhanced by allowing the user to override default-based decisions.

The expert system in Performance Instruction guides, monitors, and aggregates the decisions of an instructor in planning, delivering and evaluating instruction. The functions of the expert system are literature-based, approved practices that have been harnessed by the modern technology of a computer.

Performance Instruction holds that content included in a course should be planned, delivered, and evaluated. The integration of these elements is often neglected in traditional approaches to instruction. It is not uncommon to find no or scant planning documentation, an unfollowed syllabus, or testing that goes beyond or does not relate to what was taught.

The C-P-A model replicated in the expert system is steeped in taxonomy classifications. Vogler (1991) amalgamated the work of Bloom, et. al. (1956), Krathwohl, et. al. (1964) and Simpson (1966) in forming his own unduplicated classification scheme. For each domain, Vogler developed three classifications or levels, arranged in order from simple to complex. The matrix contains nine identifiable terms to coincide with the domain and level. For example, in the cognitive domain, the terms used to classify behavior are fact, understanding, and application. In the psychomotor domain, the descriptors used to classify expected outcomes are imitation, practice, and habit. In the affective

domain, the descriptors used to classify expected outcomes are awareness, distinction, and integration. In arriving at this condensed and less confusing classification, Vogler selected levels one, two and four from Bloom to form the boundaries of the cognitive domain. Additionally, he selected Bloom's level five and six to form the affective domain. Level three formed the psychomotor domain.

Content Goals

Careful selection of content goal action verbs produces content communication across the planning, delivery, and evaluation phases of instruction. In Performance Instruction, a present tense verb such as "define" is used in the planning phase. The verb can be changed to the gerund form "defining" to serve as the instructional topic in delivering the instruction. Finally, the same verb can be changed to the singular form "defines" and the evaluating paradigm is created.

Content goals are the communication devices that specify the outcomes of Performance Instruction. The syntax of a content goal is controlled to maximize clarity and to provide the basis for aggregating information. The expert system relies on the formatted syntax for writing content goals to help the user manipulate simple decisions to form syllabi with internal consistency and external appeal as communicative devices.

The final format of a content goal is a sentence that is typically five to nine words in length. For example, a content goal from an introductory economics course might appear as follows:

The student will identify equilibrium points. Here, the present tense action verb "identify" is linked to a direct object that may be modified by up to four adjectives. The verb will later be classified as cognitive and at a factual level. The choice of the verb, object and adjectives sets the stage for delivery and evaluation.

Through guided steps in Performance Instruction, the user identifies the domain as either cognitive, psychomotor or affective. Within each domain, there are three levels moving from simple to complex. Content with a higher level will require more time to teach and to learn. Domain and level are linked to the verb within the content goal. As a consequence, the expert system recommends the decision for the user. The decisions made at this stage affect delivery and evaluation decisions.

Verbs used in writing content goals fall into certain categories where they can be matched to one of three decision levels. Table 1 shows three levels used in planning cognitive content, arranged in ascending order of difficulty and complexity.

Table 1

Performance Instruction Domain Level Codes

Cognitive Level		Cognitive Level Code
Factual	(Level One)	1
Understanding	(Level Two)	2
Application	(Level Three)	3

The levels used by Vogler in the cognitive domain are far less confusing than having to select verbs from among six categories using Bloom's classification. In the cognitive domain, Vogler's three sub categories or levels are "fact," "understanding," and "application." These levels also create an association that implicitly relates to testing. Two other domains, psychomotor and affective, each have three sub categories or levels. Faculty select appropriate verbs pertaining to each of the levels within the appropriate domain to plan instruction. The verb is used to plan instruction by creating an outcome-based content goal statement.

The decision levels are important for course planning and testing. The levels selected by faculty when planning become important for selecting appropriate test types. Just as there are recommended verbs used at one of three cognitive planning decision levels, there are three corresponding levels used to test content.

The type of design Vogler promoted in Performance Instruction relates to similar beliefs about planning and testing held by Gronlund (1981, 1985), Diamond (1989), Nitko (1989), and Popham (1978). The domain-referenced verb in Vogler's (1991) model directs the instruction. If a verb is selected to plan instruction at the application level, the same verb should also be used when delivering and evaluating instruction to maintain consistency at the application level.

Building an Exam

The expert system in the software assures consistency between

planning and testing by recommending test types that coincide with the domain level used in the planning stage. The verbs from various taxonomies are coupled with alternative objects to form a hierarchical set of stem templates. These same verbs are linked to six test item types so that true-false and completion items will normally be used with low level cognitive content goals; matching and short answer will normally be used with middle level cognitive content goals; and multiple-choice and essay items will normally be used with high level cognitive content goals. The software recommends the type of test items based upon the domain and level of the verb in the content goal.

Table of Test Type Specifications

A table of test type specifications equates test types with domain-referenced content plans. According to Vogler (1991), test types can be paired by cognitive domain level. For example, at the factual level, the lowest level of cognitive content, the only two test types having absolute platform unity are true-false and completion items. At the understanding level, the second level, the only two test types having absolute platform unity are matching and short answer. At the application level, the highest level of cognitive content, the only two test types having absolute platform unity are multiple choice and essay. Table 2 shows the cognitive domain levels in parity with expert system test types.

Table 2

Table of Test Type Specifications

	Domain Level	Test Items
1	Fact	True-False Completion
2	Understanding	Matching Short Answer
3	Application	Multiple Choice Essay

Preparing test items directly relevant to the instructional objectives is primarily a matter of matching the performance measured by the test items to the types of performance specified by the intended outcomes. If the intended learning outcomes call for identifying a procedure, the test items should be concerned only with the process of identifying, rather than more complex outcomes. If the intended learning outcomes call for performing a procedure, the test items should require actual performance, rather than a verbal description of how to do it.

Summary

The desired outcome of any teaching strategy ought to be effective and efficient instruction. This review presented concepts deemed essential in facilitating efficiency and effectiveness of planning and evaluating from a theoretical and practical perspective. In addition, platform unity, based on the Principles of Performance Instruction, was presented as a way to uphold curriculum alignment principles. Alignment is rooted in the belief that instructional plans are established through an outcome-based content goal and assuring that the delivery and assessment are congruent.

The integration of planning and evaluating is often neglected in traditional approaches to instruction. For example, content is often planned, but not evaluated. The inverse is also true. Performance Instruction holds that course content should be planned, delivered, and evaluated consistently to assure unity.

Faculty, when planning a course, plan outcomes of instruction. The key to successful planning is to create performance-based objectives by selecting domain-referenced verbs. The same verbs are used in planning and testing to assure consistency. Testing is not a function separate from planning. When creating tests, faculty need to relate to content planning decisions. The domain-level at which content was planned becomes the basis for creating test items. According to Performance Instruction, content planning and testing occur at the same domain level to assure unity.

In the theoretical literature, considerable attention was

given to faculty's inability to plan and test content consistently. These are arguably important and necessary faculty tasks. If instruction is to be effective, then staff development efforts to explain effective planning and testing techniques should be undertaken. Unfortunately, even the literature on planning and testing treats them as separate, independent activities. Tools to assist faculty in assessing criterion-referenced tests were deemed necessary based on the lack of such measurements in the extant literature.

Faculty training in assessment ranges from very little (Coffman, 1983; Ward, 1982) to none (Stiggins & Bridgeford, 1985). Many faculty have had no formal coursework and teacher preparation programs do not require measurement training. In addition, faculty have had no in-service training in assessment (Stiggins, Conklin, & Bridgeford, 1986). Current literature suggests: (a) testing is conducted by teachers whose formal knowledge of assessment is minimal, (b) classroom tests are usually short and objective, (c) teacher-developed tests are of poor technical quality, and (d) tests generally call for memorization of facts.

In-service workshops may be one of the best ways to provide faculty with exposure to measurement techniques. Goslin (1967) reported that teachers with measurement training and experience generally tended to make the most use of tests. Kellaghan, Madaus, and Airasian (1982) argued it was not necessary for faculty to know more about interpreting test scores. Rather, they suggested offering training sessions that provide guidance in practical,

quality control measures for teachers.

There is little research in the field of tests and measurement regarding practical tools to help faculty evaluate criterion-referenced tests. The lack of faculty skills to judge the worth of teacher-made, criterion-referenced tests is apparent. The gap affects instructional practices as the use of classroom assessment techniques are increasing in postsecondary education. Standard measurement tools used to evaluate norm-referenced tests, including reliability, validity, difficulty, and discrimination are generally not applied to criterion-referenced testing situations by faculty. By spanning the vast compendium of theory related to planning and evaluation in this paper, one may see the need to move to practical implications for the benefit of aligning the planned and tested curricula thus achieving efficiency, effectiveness and overall unity of instruction.

REFERENCES

Adams, G. S. (1966). Measurement and evaluation in education, psychology, and guidance. New York: Holt, Reinhart and Winston.

Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. Journal of Educational Measurement, 20, 103-118.

Anderson, S. B. (1987). The role of teacher-made tests in higher education. In D. Bray and M. Belcher (Eds.), Issues in Student Assessment. New Directions for Community Colleges, No. 59 (pp. 39-44). San Francisco: Jossey-Bass.

Babbin, P. (1987). Instructional objectives. (ERIC Document Reproduction Service, No. ED 299 236).

Baker, E. L., & Popham, W. J. (1970). Systematic instruction. Englewood Cliffs, NJ: Prentice-Hall.

Ball, D. W., Doss, A. R., & Dewalt, M . W. (1986). Level of teacher objectives and their classroom tests: Match or mismatch. Journal of Social Studies Research, 2(10), 27-30.

Bergquist, W. H., & Armstrong, J. L. (1986). Planning effectively for educational quality. San Francisco: Jossey-Bass.

Bloom, B. S. (1961). Quality control in education. Tomorrow's teaching. Oklahoma City: Frontiers of Science Foundation.

Bloom, B. S., Englehart, M. D., Furst, G. J., Hill, W. H., & Krathwohl, D. L. (1956). Taxonomy of educational objectives: Handbook I, Cognitive domain. New York: David McKay Company.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). Evaluation to improve learning. New York: McGraw-Hill.

Boersma, W. C. (1967). The effectiveness of the evaluative criteria as a stimulus for school improvement in eleven Michigan high schools. Unpublished doctoral dissertation, University of Michigan, Ann Arbor, MI.

Brickell, H. M. (1976). Needed: Instruments as good as our eyes. Evaluation Center Occasional Paper Series No. 7. Kalamazoo, MI: Western Michigan University.

Coffman, W. E. (1983, July). Testing in the schools: A historical perspective. Paper presented at the Center for the Study of Evaluation Annual Invitational Conference, University of California at Los Angeles, Los Angeles, CA. (ERIC Document Reproduction Service No. ED 338 670)

Cook, J. M. (1978). Developing learning outcomes: Module 2 of a handbook on clarifying college learning outcomes. Columbia, MD: Council for the Advancement of Experiential Learning.

Cook, W. W. (1951). The functions of measurement in the facilitation of learning. In E. F. Lindquist (Ed.), Educational measurement. Washington, DC: American Council on Education.

Copperud, C. (1979). The test design handbook. Englewood Cliffs, NJ: Educational Technology Publications.

Cross, K. P. (1987). Teaching for learning. AAHE Bulletin, 39(8), 3-7.

Crowell, R., & Tissot, P. (1986). Curriculum alignment. Washington, DC: U. S. Department of Education. (ERIC Document Reproduction Service No. ED 280 874)

Cunningham, G. (1986). Educational and psychological measurement. New York: Macmillan.

Diamond, R. M. (1989). Designing and improving courses and curricula in higher education: A systematic approach. San Francisco: Jossey-Bass.

Dyer, H. S. (1970). Basic elements of the measurement process. In Norman E. Gronlund (Ed.), Readings in measurement and evaluation (pp. 3-10). New York: Macmillan.

Ebel, R. L. (1965). Measuring educational achievement. Englewood Cliffs, NJ: Prentice-Hall.

Ebel, R. L., & Frisbie, D. (1986). Essentials of educational measurement (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

English, F. W. (1979). Management practice as key to curriculum leadership. Educational Leadership, 36, 408-413.

English, F. W. (1986-1987). It's time to abolish conventional curriculum guides. Educational Leadership, 44(4), 50-53.

English, F. W. (1988). Curriculum auditing. Lancaster, PA: Technomic Publishing Co.

English, F. W. (Ed.). (1992a, January). Curriculum audit training handbook. (Available from AASA, 1801 N. Moore St., Arlington, VA 22209).

English, F. W. (1992b). Deciding what to teach and test: Developing, aligning, and auditing the curriculum. Newberry Park, CA: Corwin Press.

English, F. W., & Steffy, B. E. (1983). Differentiating between design and delivery problems in achieving quality control in school curriculum management. Educational Technology, 24(2) 29-32.

Feldhausen, J. F. (1980). The three-stage model of course design. Englewood Cliffs, NJ: Educational Technology Publications.

Fleming, M., & Chambers, B. (1983). Teacher-made tests: Windows on the classroom. In W. E. Hathaway (Ed.), Testing in the schools. New Directions for Testing and Measurement, No. 19 (pp. 29-38). San Francisco: Jossey Bass.

Fortune, J. C., and Associates (1985). Understanding testing in occupational licensing. San Francisco: Jossey-Bass.

Gaff, J. G. (1975). Toward faculty renewal. San Francisco: Jossey-Bass.

Gagne, R. M. (1975). Essentials of learning for instruction. New York: Holt, Reinhart and Winston.

Gentry, D. L. (1989). Teacher-made test construction. (ERIC Document Reproduction Service No. ED 313 444).

Gerhard, M. (1971). Effective teaching strategies with the behavioral outcomes approach. West Nyack, NY: Parker Publishing Company.

Goslin, D. A. (1967). Teachers and testing. New York: Russell Sage Foundation.

Green, J. A. (1975). Teacher-made tests (2nd ed.). New York: Harper & Row.

Greeno, J. G. (1976). Cognitive objectives of instruction: Theory of knowledge for solving problems and answering questions. In David Klahr (Ed.), Cognition and instruction (pp. 3-11). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gronlund, N. E. (1981). Measurement and evaluation in teaching. (4th ed.). New York: Macmillan.

Gronlund, N. E. (1985). Stating objectives for classroom instruction (3rd ed.). New York: Macmillan.

Gunter, M. A., Estes, T. H., & Schwab, J. H. (1990). Instruction: A models approach. Boston: Allyn and Bacon.

Hartzell, M. S. (1984). Checking for curriculum/test overlap: Two methods discussed. Portland, OR: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED 246 081).

Heywood, J. (1989). Assessment in higher education (2nd ed.). New York: John Wiley and Sons.

Isaac, S., & Michael, W. B. (1990). Handbook in research and evaluation. (2nd ed.). San Diego: EdITS.

Jacobs, L. C., & Chase, C. I. (1992). Developing and using tests effectively: A guide for faculty. San Francisco: Jossey-Bass.

Kellaghan, T., Madaus, G. F., & Airasian, P. W. (1982). The effects of standardized testing. Boston: Kluwer-Nijhoff.

Kibler, R. J., Cegala, D. J., Watson, K. W., Barker, L. L., & Miles, D. T. (1981). Objectives for instruction and evaluation (2nd ed.). Boston: Allyn and Bacon.

Krathwohl, D. L., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives: Handbook II, Affective domain. New York: David McKay Company.

Kropp, R. P., Stoker, H. W., & Bashaw, W. L. (1966). The construction and validation of tests of the cognitive processes as described in the taxonomy of educational objectives. Cooperative Research Project No. 2117. Washington, DC: Office of Education, U. S. Department of Health, Education, and Welfare.

Leinhardt, G., & Seewald, A. M. (1981). Overlap: What's tested, what's taught? Journal of Educational Measurement, 18, 85-95.

Leitzel, T. C., & Vogler, D. E. (1993). Platform unity/alignment between course planning and testing decisions in criterion-referenced situations. (Doctoral dissertation, Virginia Polytechnic Institute and State University, 1993). Ann Arbor, MI: University Microfilms International No. 9319779

Linn, R. L. (1983). Testing and Instruction: Links and distinctions. Journal of Educational Measurement, 20, 179-189.

Linn, R. L. (1989). Current perspectives and future directions. In Robert L. Linn (Ed.) Educational measurement. (3rd ed.) (pp. 1-10). New York: Macmillan.

Lozak, J. (1987). Assessment and improvement in education. In D. Bray and M. Belcher (Eds.), Issues in Student Assessment. New Directions for Community Colleges, No. 59 (pp. 25-30). San Francisco: Jossey-Bass.

MacCuish, D. A. (1986, February). A quality assurance model for higher education: A pilot study. paper presented at the Testing and Quality Assurance in Higher Education Conference. (ERIC Document Reproduction Service No. ED 273 168)

Mager, R. F. (1962). Preparing instructional objectives. San Francisco: Fearon.

Mager, R. F. (1973). Measuring instructional intent. Belmont, CA: Fearon-Pittman.

Mehrens, W. A., & Lehmann, I. J. (1984). Measurement and evaluation in educational psychology. (3rd ed.). New York: Holt, Reinhart & Winston.

Menges, R. J., & Mathis, C. B. (1988). Key resources on teaching, learning, curriculum, and faculty development. San Francisco: Jossey-Bass.

Metfessel, N. S., Michael, W. B., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the writing of educational objectives. Psychology in the Schools, 6, pp. 227-231.

Milton, O., and Associates (1978). On college teaching: A guide to contemporary practices. San Francisco: Jossey-Bass.

Nitko, A. J. (1989). Designing tests that are integrated with instruction. In Robert L. Linn (Ed.), Educational measurement. (3rd ed.) (pp. 447-474). New York: Macmillan.

Pautler, A. J. (1989). Curriculum alignment as a means for improving institution effectiveness. Community/Junior College Quarterly of Research and Practice, 13, 173-180.

Pautler, A. J. (1990). Designing functional courses of study for community colleges. Journal of Studies in Technical Careers, 12, 173-181.

Pautler, A. J. (1992). Curriculum leadership, innovation, and change. Community/Junior College Quarterly of Research and Practice, 16, 133-140.

Payne, D. A. (1974). The assessment of learning. Lexington, MA: D. C. Heath.

Payne, D. A. (1992). Measuring and evaluating educational outcomes. New York: Macmillan.

Popham, W. J. (1973). The uses of instructional objectives: A personal perspective. Belmont, CA: Fearon.

Popham, W. J. (1978). Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall.

Powell, J. L., & Gillespie, C. (1990, December). Assessment: all tests are not created equally. Paper presented at the Annual Meeting of the American Reading Forum, Sarasota, FL. (ERIC Document Reproduction Service No. ED 328 908)

Roid, G., & Haladyna, T. (1982). A technology for test item writing. New York: Academic.

Salvia, J., & Hughes, C. (1990). Curriculum-based assessment: Testing what is taught. New York: Macmillan.

Savard, W. G., & Cotton, K. (1982). Curriculum alignment: Research on school effectiveness project. (ERIC Document Reproduction Service No. ED 265 631)

Scott, R. (1983). Curriculum alignment as a model for school improvement. (ERIC Document Reproduction Service No. ED 252 508).

Simpson, E. J. (1966). The classification of educational objectives: Psychomotor domain. Urbana: University of Illinois Press.

Stahl, R. J., & Murphy, G. T. (1981, April). The domain of cognition: An alternative to Bloom's cognitive domain within the framework of an information processing model. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles. (ERIC Document Reproduction Service No. ED 314 854).

Stanley, J. C., & Bolton, D. L. (1957). A review of Bloom's taxonomy of educational objectives. Educational and Psychological Measurement, 17, 631-634.

Stanley, J. C., & Hopkins, K. D. (1972) Educational and psychological measurement and evaluation. Englewood Cliffs, NJ: Prentice-Hall.

Stark, J. S., Lowther, M. A., Sossen, P. L. & Shaw, K. M. (1991). Course planning exploration for program self-study. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.

Stiggins, R. J., & Bridgeford, N. J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22, 271-286.

Stiggins, R. J., Conklin, N. F., & Bridgeford, N. J. (1986). Classroom assessment: A key to effective instruction. Educational measurement: Issues and Practice, 5(2) 5-17.

Swezey, R. (1981). Individual performance assessment: An approach to criterion-referenced test development. Reston, VA: Reston.

Thorndike, R., & Hagen, E. (1969). Measurement and evaluation in psychology and education (3rd ed.). New York: John Wiley and Sons.

Tuckman, B. W. (1975). Measuring educational outcomes: Fundamentals of testing. New York: Harcourt Brace Jovanovich.

Tyler, R. W. (1951). The functions of measurement in improving instruction. In E. F. Lindquist (Ed.), Educational measurement. Washington, DC: American Council on Education.

Vogler, D. E. (1991). Performance instruction: Planning, delivering, evaluating. Eden Prairie, MN: Instructional Performance Systems, Inc.

Ward, J. G. (1982). An overview of the AFT's teaching and testing. In S. B. Anderson and L. B. Colburn (Eds.) Academic testing and the consumer (pp. 47-52). San Francisco: Jossey-Bass.

Wesman, A. (1971). Writing the test item. In R. Thorndike (Ed.) Educational Measurement (2nd ed.) (pp. 81-130). Washington, DC: American Council on Education.

Yeh, J. P. (1978). Test use in schools. Washington, DC: U. S. Department of Health, Education and Welfare and National Institute of Education.